Appendix B. Source and Reliability of the Estimates

SOURCE OF DATA

The estimates of vocational school experience in 1976 are based on data obtained in October of 1976 in the Current Population Survey (CPS) conducted by the Bureau of the Census. The CPS sample was initially selected from the 1970 census file and is updated continuously to reflect new construction where possible. The current sample is spread over 461 areas with coverage in each of the 50 States and the District of Columbia. In the sample, approximately 47,000 households are eligible for interview each month. Of this number, 2,000 occupied housing units on the average, are visited but interviews are not obtained because the occupants are not found at home after repeated calls or are unavailable for some other reason. In addition to the 47,000, there are also about 8,000 sample units in an average month which are visited but are found to be vacant or otherwise not to be interviewed.

The estimating procedure used for the monthly CPS data involves the inflation of the weighted sample results to independent estimates of the total civilian noninstitutional population of the United States by age, race, and sex. These independent estimates are based on statistics from decennial censuses; statistics on births, deaths, immigration and emigration; and statistics on the strength of the Armed Forces.

RELIABILITY OF THE ESTIMATES

Since the CPS estimates in this report were based on a sample, they may differ somewhat from the figures that would have been obtained if a complete census had been taken using the same questionnaires, instructions, and enumerators. There are two types of errors possible in an estimate based on a sample survey—sampling and nonsampling. The standard errors provided for this report primarily indicate the magnitude of the sampling errors. They also partially measure the effect of some nonsampling errors in response and enumeration, but do not measure any systematic biases in the data. The full extent of the nonsampling error is unknown. Consequently, particular care should be exercised in the interpretation of figures based on a relatively small number of cases or on small differences between estimates.

Nonsampling variability. Nonsampling errors can be attributed to many sources, e.g., inability to obtain information about all cases in the sample, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, inability to recall information, errors made in collection such as in recording or coding the data, errors made in processing the data, errors made in estimating values for missing data, and failure to represent all units with the sample (undercoverage).

Undercoverage in the CPS results from missed housing units and missed persons within sample households. Overall undercoverage, as compared to the level of the decennial census, is about 5 percent. It is known that CPS undercoverage varies with age, sex, and race. Generally, undercoverage is larger for males than for females and larger for Blacks and other races than for Whites. Ratio estimation to independent age-sex-race population controls, as described previously, partially corrects for the bias due to survey undercoverage. However, biases exist in the estimates to the extent that missed persons in missed households or missed persons in interviewed households have different characteristics than interviewed persons in the same age-sex-race group. Further, the independent population controls used have not been adjusted for undercoverage in the 1970 census, which was estimated at 2.5 percent of the population, with similar undercoverage differentials by age, sex, and race as in CPS.

The approximate magnitude of two sources of undercoverage of housing units is known. Of the 83,000,000 housing units in the U.S., about 600,000 new construction housing units other than mobile homes are not represented in the CPS sample because they were assigned building permits prior to January 1970, but building was not completed by the time of the census (i.e., April 1970). Almost all conventional new construction, for which building permits were issued after 1969, is represented. About 290,000 occupied mobile homes are not represented in CPS; these units were either missed in the census or have been built or occupied since the census. These estimates of missed units are relevant to the present sample only and not to earlier designs where the extent of undercoverage was generally less. The extent of other sources of undercoverage of housing units is unknown but believed to be small.

Sampling variability. The standard errors given in the following tables are primarily measures of sampling variability, that is, of the variations that occurred by chance because a sample rather than the entire population was surveyed. The sample estimate and its estimated error enable one to construct confidence intervals, ranges that would include the average result of all possible samples with a known probability. For example, if all possible samples were

selected, each of these surveyed under essentially the same general conditions and using the same sample design, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average estimate derived from all possible samples is or is not contained in any particular computed interval. However, for a particular sample, one can say with a specified confidence that the average estimate derived from all possible samples is included in the confidence interval.

All the statements of comparison appearing in the text are significant at a 1.6 standard error level or better, and most are significant at a level of more than 2.0 standard errors. This means that for most differences cited in the text, the estimated difference is greater than twice the standard error of the difference. Statements of comparison qualified in some way (e.g., by use of the phrase, "some evidence") have a level of significance between 1.6 and 2.0 standard errors.

Comparability with other data. Caution should be used when comparing income data from this report with the March Supplemental Income Data. For a further discussion of this see the definition and explanation for family income in Appendix A of this report.

Note when using small estimates. Summary measures (such as percent distributions) are shown in the report only when the base of the measure is 75,000 or greater. Because of the large standard errors involved, there is little chance that summary measures would reveal useful information when computed on a smaller base. Estimated numbers are shown, however, even though the relative standard errors of these numbers are larger than those for corresponding percentages. These smaller estimates are provided primarily to permit such combinations of the categories as serve each user's need.

Standard error tables and their use. In order to derive standard errors that would be applicable to a large number of estimates and could be prepared at a moderate cost, a number of approximations were required. Therefore, instead of providing an individual standard error for each estimate, generalized sets of standard errors are provided for various types of characteristics. As a result, the sets of standard errors provided give an indication of the order of magnitude of the standard error of an estimate rather than the precise standard error.

The figures in tables B-1 and B-2 provide approximations to standard errors of estimated numbers and estimated percentages. Standard errors for intermediate values not shown in the generalized tables of standard errors may be approximated by linear interpolation. Estimated standard errors for specific characteristics cannot be obtained from tables B-1 or B-2 without the use of factors in table B-3. These factors must be applied to the generalized standard errors in order to adjust for the combined effect of sample design and estimating procedure on the value of the characteristic.

Two parameters (denoted "a" and "b") are used to calculate standard errors for each type of characteristic; they are presented in table B-3. These parameters were used to calculate the standard errors in tables B-1 and B-2, and to calculate the factors in table B-3. They also may be used to directly calculate the standard errors for estimated numbers and percentages. Methods for direct computation are given in the following sections.

Standard errors of estimated numbers. The approximate standard error, σ_x, of an estimated number shown in this report can be obtained in two ways. It may be obtained by use of the formula

$$\sigma_x = f\sigma \qquad (1)$$

where f is the appropriate factor from table B-3, and σ is the standard error on the estimate obtained by interpolation from table B-1. Alternatively, standard errors may be approximated by formula (2) from which the standard errors were calculated in table B-1. Use of this formula will provide more accurate results than the use of formula (1) above.

$$\sigma_x = \sqrt{ax^2 + bx} \qquad (2)$$

Here x is the size of the estimate and a and b are the parameters in table B-3 associated with the particular type of characteristic.

Standard errors of estimated percentages. The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more. When the numerator and denominator of the percentage are in different categories, use the factor or parameters from table B-3 indicated by the numerator. The approximate standard error, $\sigma_{(x,p)}$, of an estimated percentage can be obtained by use of the formula

$$\sigma_{(x,p)} = f\sigma \qquad (3)$$

In this formula f is the appropriate factor from table B-3, and σ is the standard error on the estimate from table B-2. Alternatively, standard errors may be approximated by formula (4), for which standard errors in table B-2 were

calculated; direct computation will give more accurate results than use of the standard error tables and the factors.

$$\sigma_{(x,p)} = \sqrt{\frac{b}{x} \cdot p\,(100 - p)} \qquad (4)$$

Here x is the size of the subclass of persons, families and unrelated individuals, households, householders or primary individuals which is the base of the percentage, p is the percentage ($0 \leq p \leq 100$), and b is the parameter in table B-3 associated with the particular type of characteristic in the numerator of the percentage.

Illustration of the use of standard error tables. Table B of this report shows that 227,000 Blacks, 14 years and over had attended vocational school in October 1976. Table B-1 shows that the standard error on an estimate of this size to be approximately 22,000. Applying the appropriate factor from table B-3 and using formula (1), the approximate standard error is 1.16 x 22,000 = 26,000[1]. The 68-percent confidence interval as shown by the data is from 201,000 to 253,000. Therefore a conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 68 percent of all samples. Similarly we could conclude that the average estimate derived from all possible samples lies within the interval from 175,000 to 279,000 (using twice the standard error) with 95-percent confidence.

These 227,000 Blacks who had attended vocational school in October 1976 represented 41.1 percent of the 552,000 Blacks, 14 years and over with recent vocational school experience. Since the numerator of this percentage is Blacks with vocational school training, the appropriate b parameter from table B-3 is 2,792. Using formula (4), the standard error on an estimate of 41.1 percent is

$$\sqrt{\frac{2{,}792}{552{,}000}\,(41.1)\,(58.9)} \doteq 3.5 \text{ percent}^2$$

[1] Formula (2) for this example also gives a standard error of 26,000.

[2] Table B-2 and formula (3) give a standard error of 3.3 percent.

Table B-1. Standard Errors of Estimated Numbers

(Numbers in thousands)

Size of estimate	Standard error	Size of estimate	Standard error
25	7	5,000	100
50	10	10,000	138
100	14	15,000	165
250	23	25,000	204
500	32	50,000	251
1,000	45	100,000	215
2,500	71		

Note: For particular characteristic see table B-3 for the appropriate factor to apply to the above standard errors.

Table B-2. Standard Errors of Estimated Percentages

Base of estimated percentage (thousands)	Estimated percentage					
	1 or 99	2 or 98	5 or 95	10 or 90	25 or 75	50
75	1.7	2.3	3.6	5.0	7.2	8.3
100	1.4	2.0	3.1	4.3	6.2	7.2
250	0.9	1.3	2.0	2.7	3.9	4.5
500	0.6	0.9	1.4	1.9	2.8	3.2
1,000	0.5	0.6	1.0	1.4	2.0	2.3
2,500	0.3	0.4	0.6	0.9	1.2	1.4
5,000	0.2	0.3	0.4	0.6	0.9	1.0
10,000	0.14	0.2	0.3	0.4	0.6	0.7
15,000	0.12	0.2	0.3	0.4	0.5	0.6
25,000	0.09	0.13	0.2	0.3	0.4	0.5
50,000	0.06	0.09	0.14	0.2	0.3	0.3
100,000	0.05	0.06	0.10	0.14	0.2	0.2

Note: For a particular characteristic see table B-3 for the appropriate factor to apply to the above standard errors.

Consequently, the 68 percent confidence interval as shown by these data is from 37.6 to 44.6 percent, and the 95 percent confidence interval is 34.1 to 48.1 percent.

Standard error of a difference. For a difference between two sample estimates, the standard error is approximately equal to

$$\sigma_{(x-y)} = \sqrt{\sigma_x^2 + \sigma_y^2} \qquad (5)$$

where σ_x and σ_y are the standard errors of the estimates x and y; the estimates can be of numbers, percents, ratios, etc. This will represent the actual standard errors quite accurately for the difference between two estimates of the same characteristic in two different areas, or for the difference between separate and uncorrelated characteristics in the same area. If, however, there is a high positive (negative) correlation between the two characteristics, the formula will overestimate (underestimate) the true standard error.

Illustration of the computation of the standard error of a difference. Table B of this report also shows that 4.6 percent of all Whites 14 years and over had recent vocational school experience in October 1976, while 3.2 percent of all Blacks 14 years and over had recent vocational school experience. The apparent difference between the percent of Whites and Blacks with recent vocational school experience in October 1976 is 1.4 percent. The standard error on the 4.6 percent of Whites with vocational school experience is 0.08 percent. The standard error on the 3.2 percent of Blacks with vocational school experience is 0.2 percent. Therefore the standard error of the estimated difference of 1.4 percent is about

$$0.2 \text{ percent} \doteq \sqrt{(0.08)^2 + (0.2)^2}$$

This means that the 68 percent confidence interval for the difference is from 1.2 to 1.6 percent. A conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. The corresponding 95-percent confidence interval is 1.0 to 1.8 percent. Thus, we can conclude with 95-percent confidence that the percentage of Whites 14 years and over with recent vocational school experience in October 1976 is greater than the corresponding percentage of Blacks.

Table B-3. "a" and "b" Parameters and Factors for Estimated Numbers and Percentages

Type of characteristic	Parameters		Factors
	a	b	F
Vocational School Training, Field of Training and Occupational Training:			
Total or White	-0.000016	2064	1.00
Black and other races	-0.000186	2792	1.16
Spanish origin	-0.000025	3851	1.37
Employment:			
Both Sexes			
Total or White	-0.000016	2078	1.00
Black and other races	-0.000133	2078	1.00
Spanish origin	-0.000026	4394	1.46
Male			
Total or White	-0.000025	1798	0.93
Black and other races	-0.000221	1798	0.93
Female			
Total or White	-0.000019	1541	0.86
Black and other races	-0.000152	1541	0.86
Income:			
Family			
Total or White	-0.000008	1063	0.72
Black and other races	-0.000064	922	0.67
Spanish origin	-0.000033	2397	1.08

☆U.S. GOVERNMENT PRINTING OFFICE: 1979— 281-042:189